21cm cosmology meets Artificial Neural Networks

Hayato Shimabukuro [島袋隼士] (Observatoire de Paris)

Based on Shimabukuro & Semelin
(astro-ph/1701.07026, Accepted in MNRAS)

@ASIAA
Contents

• Introduction
• Cosmological 21cm signal
• Artificial neural networks (ANN)
• Result
• Summary
Introduction
Who am I?

• Okinawa-Tohoku University-Nagoya University (Kumamoto University)

• Ph.D at Nagoya university (03/2016)

• Post-doc at Paris observatory (Supervisor : Benoit Semelin, 04/2016~)

• This is second visit to Taiwan (since 12/2015)
Dark age: Star formation has not started yet.

Cosmic dawn: After star formation, astrophysical effects become effective.

Epoch of Reionization (EoR): Hydrogen is ionized by UV photons emitted from stars, galaxies.
Epoch of Reionization

Pre-overlap: HII regions grow in relative isolation

Overlap: Once galaxies become common, HII regions intersect.

Post-overlap: Ionising of IGM advances sufficiently.

Blue: Neutral IGM       Red: HII region    Yellow: Ionizing source (galaxy)
Observational constraints on EoR

† The polarisation of CMB photons by free electron.

\[ z_r = 9.9^{+1.8}_{-1.6} \]  (Assuming instantaneous reionization model)

(Planck collaboration 2015)

† High-z QSO absorption spectra

\[ x_{\text{HI}} > \sim 10^{-4} (z > 6) \]  (Put constraints on neutral fraction)

(Fan et al 2006)

† The luminosity function of high-z LAE (Lyman-alpha emitter galaxies)

\[ x_{\text{HI}} = 0.3 - 0.8 \ (z = 7.3) \]  (Put constraints on neutral fraction)

(Konno et al 2014)
Beyond current constraints

Current observations can constrain state of the IGM at **late** stage of the EoR.

**We would like to observe the higher redshift universe more directly!**

---

**21cm line signal** from neutral hydrogen in the IGM

---

**Simulated 21cm map**

- **z=12.1**
- **z=9.2**
- **z=7.6**

---

Neutral **Ionized**

---

(Furlanetto & Briggs 2004)
Cosmological 21cm signal
**21cm signal**

○ **21cm line radiation**: Neutral hydrogen emits the radiation due to the hyperfine structure.

\[ \Delta E = 5.9 \times 10^{-6} \text{eV} \]

Spin temperature

\[ \frac{n_1}{n_0} = 3 \exp \left( - \frac{h \nu_{21}}{kT_S} \right) \]

○ Collision with CMB photons

○ Collision with Ly-α photons

○ Collision with atoms
Brightness temperature

We actually observe brightness temperature instead of spin temperature.

\[ \delta T_b(\nu) = \frac{T_S - T_\gamma}{1 + \frac{1}{z}} (1 - e^{-\tau_{\nu_0}}) \]

\[ \sim 27 x_H (1 + \delta_m) \left( \frac{H}{dv_r/dr + H} \right) \left(1 - \frac{T_\gamma}{T_S} \right) \left(1 + \frac{z}{10} \frac{0.15}{\Omega_m h^2} \right)^{1/2} \left( \frac{\Omega_b h^2}{0.023} \right)[\text{mK}] \]

Red: astrophysics.
Blue: cosmology

Pritchard & Loeb (2011)
21cm power spectrum

We often evaluate statistical property of the 21cm fluctuations by power spectrum.

21cm power spectrum (PS)

$$\langle \delta T_b(\mathbf{k})\delta T_b(\mathbf{k}') \rangle = (2\pi)^3 \delta(\mathbf{k} + \mathbf{k}') P_{21}$$

**Scale dependence**

![Graph showing scale dependence](image)

**Redshift dependence**

![Graph showing redshift dependence](image)

Pober et al (2014)
Observations

Some on-going telescopes have started observation (MWA, LOFAR, PAPER). Future observations are now planning (SKA, HERA) on 2020’s.

Collaboration with SKA-JAPAN has started (Kumamoto, Nagoya, ICRR).

Cross-correlation between 21cm - LAE (Kubota et al, in preparation)
Upper limits on 21cm PS


Current upper limit by PAPER

- 1-2 magnitude of order higher than theoretical prediction.
- But, we believe 21cm signal is detectable in near future.

After we actually detect 21cm signal.

How do we extract astrophysical information from 21cm PS?

Theoretical prediction

Artificial neural network
Artificial neural networks
Motivation

• We would like to extract the EoR information from 21cm PS.
• To determine EoR parameters helps this purpose.
• How precisely can we determine EoR model parameters?
• We train neural network architecture to learn association between 21cm PS and EoR parameters.
• Once we train ANN architecture, we can apply this to unknown data.
What is artificial neural network (ANN)?

ANN is one of the methods inspired by brain neural network which is used to establish approximate function between input and output data.

STEP

1. Prepare known data set (training data) \( (x_{data}, y_{data}) \)

2. Train architecture of neural network by training data.
   \[ y = f(x) \]

3. Apply trained network to unknown (test data) and can obtain expected output data.
   \[ y_{ANN} = f(x_{test}) \]
Flowchart

1. Training set
2. Machine learning algorithm
3. Test data
4. Trained neural network

Predicted output
What is ANN?

1. At input layer, we first calculate the linear combination of input data with weight and transfer them to hidden layer.

2. Next, we activate that linear combination with activation function.

\[
\text{Neuron}\quad s_i = \sum_{j=1}^{n} w_{ij} x_j \\
\phi(s_i) = t_i
\]

\(w_{ij}:\text{Weight}\)

\(\phi: \text{Activation function}\)
What is ANN?

At output layer, we consider linear combination of neurons at hidden layer. Unlike hidden layer, we do not need to activate linear combination at output layer.
What is ANN?

We have to determine the weight to construct architecture of neural network.

**Answer**

We adjust the weight to satisfy minimizing the cost function $E$ which is difference between true value and output value. This procedure is called “Training”

**cost function**

$$E = \frac{1}{2} \sum_{i}^{N} (y_{data,i} - y_{ANN})^2$$

$N$: the number of training datasets

We use “Back propagation algorithm” to determine weight. (Rumelhart et al 1986)
Back-propagation algorithm

We update the weights by gradient descent of cost function until they converge.

\[ w(t + 1) = w(t) + \Delta w(t) \]

with

\[ \Delta w^{(l)}_{ij} = -\eta \frac{\partial E}{\partial w^{(l)}_{ij}} = \eta \sum_{n=1}^{N_{\text{train}}} \frac{\partial E_n}{\partial w^{(l)}_{ij}} \]

We calculate the derivative of cost function starting from output layer toward input layer.
Dataset

\[ \vec{d} = [P(k), \vec{\theta}] \]

\[ \theta_{\text{EoR}} = f(P_{21}) \]

EoR parameter (output)

21cm power spectrum (input) \((0.04\text{Mpc}^{-1} \leq k \leq 1.4\text{Mpc}^{-1})\)

(14 bins)

EoR Parameter

\( \zeta \): the ionizing efficiency.

\( T_{\text{vir}} \): the minimum virial temperature of halos producing ionizing photons

\( R_{\text{mfp}} \): the mean free path of ionizing photons through the IGM

(Maximum HII bubble size)
We used 70 training datasets and 50 test datasets.

We calculate the 21cm PS by 21cmFAST.

We choose EoR parameters and 21cm PS datasets at $z=9,12$ for single $z$ (and $z=9,10,11$ for multiple $z$).

We train network with/without thermal noise and cosmic variance.

$N_{\text{input}}=14, N_{\text{hidden}}=14, N_{\text{output}}=3$ for single $z$. 
Result
We perform $10^6$ iterations for back-propagation algorithm to see convergence.

![Graph showing convergence of RMSE over iterations]

$$RMSE = \sqrt{\frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} X^2}$$

$$X = \frac{(\theta_{\text{ANN}} - \theta_{\text{data}})}{\theta_{\text{data}}}$$

($\theta$ is each EoR parameter.)

More than 200000 iterations, all values of RMSE converge.
The number of neurons

- Fix 100000 iterations
- Change the number of neurons at hidden layer.
- Plot RMSE as function of the number of neurons.

(Note)
14 neurons at input layer.

Week dependence on the number of neurons.
$z=12$, PS without any noise

14 neurons, 100000 iterations

- True value vs. Reconstructed value
- The scatter of $R_{mfp}$ is large.
- Other reconstructed parameters match true one relatively well.
Results

$z=9$, PS without any noise

14 neurons, 100000 iterations

- Compared with $z=12$, reconstructed $R_{\text{mfp}}$ match true value better.

- Because $R_{\text{mfp}}$ expresses maximum size of HII bubble, it affects lower redshift when reionization advances.
In order to evaluate the effect of noise, we include both thermal noise and cosmic variance for test data. However, we include only cosmic variance for training data.

How?

• We calculate cosmic variance from 10 realizations of initial condition.

• Produce 50 noise realizations for each parameter set by adding random gaussian noise.

We calculate thermal noise assuming SKA observation based on (Morales et al 2005).
Results

\( z=9 \), PS including thermal noise and cosmic variance

- **Accuracy is better when noises are not included.**

<table>
<thead>
<tr>
<th></th>
<th>( R_{\text{mfp, true}} ) [Mpc]</th>
<th>( \zeta_{\text{true}} )</th>
<th>( T_{\text{vir, true}} [K/10^3] )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RMSE_wo/_noise</strong></td>
<td>0.228</td>
<td>0.271</td>
<td>0.027</td>
</tr>
<tr>
<td><strong>RMSE_w/_noise</strong></td>
<td>0.258</td>
<td>0.288</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Results

*z*=9, 10, 11. PS including thermal noise and cosmic variance

Visually, it is difficult to see whether the accuracy of parameter estimation is improved by taking redshift evolution into account.
Results

Reduce the number of training datasets (N=20).

- Estimate of the EoR parameter is improved by
  - Taking redshift evolution into account.
  - Increasing the number of training datasets
Summary

- Artificial neural network (ANN) is one of the machine learning techniques based on brain architecture model.
- We applied the ANN to 21cm signal in order to extract EoR information.
- EoR parameters produced by ANN were good agreement with true values.
- Multiple redshift data improved accuracy.
- We’re now trying to analyse the epoch of cosmic dawn with more parameters (with A. Fialkov).
Backup
Compare with MCMC

<table>
<thead>
<tr>
<th></th>
<th>ANN $\text{RMSE}_{\text{SKA}}$</th>
<th>ANN $\text{RMSE}_{\text{HERA}}$</th>
<th>MCMC $1\sigma_{\text{SKA}}$</th>
<th>MCMC $1\sigma_{\text{HERA}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{\text{mfp}}$</td>
<td>0.258</td>
<td>0.278</td>
<td>0.178</td>
<td>0.184</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>0.288</td>
<td>0.354</td>
<td>0.167</td>
<td>0.220</td>
</tr>
<tr>
<td>log($T_{\text{vir}}$)</td>
<td>0.038</td>
<td>0.040</td>
<td>0.024</td>
<td>0.033</td>
</tr>
</tbody>
</table>

- For comparison with MCMC, we also show 1 sigma error obtained by MCMC.

- We compare the RMSE obtained by ANN with 1 sigma error obtained by MCMC in the both case of HERA and SKA observation.

- EoR parameters obtained by ANN have similar error level to those by MCMC.
Thermal history

Mesinger et al 2010

\[ T_S^{-1} = \frac{T_{\text{CMB}}^{-1} + x_\alpha T^{-1} + x_K T_K^{-1}}{1 + x_\alpha + x_K} \]

**Blue**: CMB temperature
\[ \propto (1 + z) \]

**Green**: kinetic temperature
\[ \propto (1 + z)^2 \]

**Red**: spin temperature
Thermal history

Mesinger et al 2010

\[ T_{S}^{-1} = \frac{T_{\text{CMB}}^{-1} + x_{\alpha}T_{\alpha}^{-1} + x_{K}T_{K}^{-1}}{1 + x_{\alpha} + x_{K}} \]

X-ray heating

collision coupling

Spin temperature couples to kinetic temperature via Ly-\(\alpha\) photons.
Constraints

The observation of the CMB polarization
→ the optical depth of Thomson scattering

Optical depth

$$\tau_e \propto \int_{z_r}^{0} n_e(z) \frac{dt}{dz} dz$$

$$\tau_e = 0.078 \pm 0.0019$$

$$z_r = 9.9^{+1.8}_{-1.6}$$

(Planck collaboration 2015)
Constraints

high-z QSO absorption

→ constraint on the epoch where the EoR finishes.

**Gun-Peterson test**

If the neutral hydrogen exists, it absorbs the Ly-alpha photons. **No emission line!**

We can know the epoch where EoR finished via the QSO spectrum.
Sensitivity

![Sensitivity Graph](image)
Results

Reduce the number of training data. \((N=20)\)
The probability distribution function of each parameter.
Activation function

\textit{tanh}(x)

\textit{Sigmoid}, \((1 + \exp(-x))^{-1}\)

\textit{Hard Limiter}

\textit{Ramp Function}